UNITED
NATIONS

E

Economic and Social Council

Distr.
LIMITED

E/CONF.85/L.80
22 July 1992

ENGLISH ONLY

SIXTH UNITED NATIONS CONFERENCE
ON THE STANDARDIZATION OF
GEOGRAPHICAL NAMES
New York, 25 August-3 September 1992
Item 11 (a) of the provisional agenda*

WRITING SYSTEMS AND GUIDES TO PRONUNCIATION: ROMANIZATION

The romanized transliteration of Thai

Paper submitted by Thailand**

---

* E/CONF.85/1/Rev.1.

** Prepared by Prasong Chingchai, Royal Thai Survey Department, Thailand.

92-33439 3390j (E) 030892 *050892*

/...

1. INTRODUCTION

Thailand has presented the paper on the Romanized Transliteration of Thai as the information paper at the Fifth United Nations Conference on the Standardization of Geographical Names and as the technical paper at the Fifteen Session of the United Nations Group of Experts on Geographical Names as shown in document E/CONF.79/INF/39 and Working Paper No. 41 respectively. In those papers explaining that the Royal Institute set up the Thai Romanization Improvement Committee to handle the improvement of Romanization because of the need to have a new Romanization for automated information or data exchange and also need to modified the existing system for more practical use. The Committee has finally came up with the two seperate topics : Romanized Transliteration System and Romanized Transcription System. This paper will emphasize only on the first one, i.e. the Romanized Transliteration System or Transliteration of Thai. This system has already been adopted as DIS by the International Standardization Organization (ISO). Thailand submitted this paper to the Uited Nations Conference to consider and accepted for the purpose of international data exchange by the computerized system.

2. SCOPE AND FIELD OF APPLICATION

The Romanized transliteration system is a means of converting the Thai writing system into Roman alphabet. Since there are fewer Roman characters than Thai characters, diacritical marks, punctuation marks, and

a combination of two Roman characters, or a diagraph, are needed to represent one Thai character. The aim of this system is to provide a means for international communication of written messages in a form which permits the automatic transmission or reconstitution of these by men or machines. It can be seen that the transcription is suitable for human use where as transliteration is suitable for machine use. This system of conversion is intended to provide complete and unambiguous reversibility. It is possible that the results obtained from this system will not provide for correct pronunciation of the original Thai text, but it will serve as a means of finding automatically the Thai graphism and thus to allow anyone with a knowledge of Thai to pronounce the Thai text correctly. An attempt to transpose a preposed vowel after an initial consonant may comfort some who are used to the Roman writing system where a vowel always follows a consonant to form a syllable unless it is a syllable without an initial. It is best leaving the preposed vowel in the Thai transliteration system where it is because a transliteration system is not a transcription system.

Capital letters are reserved for writing an initial in proper nouns. It is unwise to make use capital letters in the transliteration system for certain consonants; it has been found out that common nouns and proper nouns cannot be distinguished from one another.

This system is based on phonetic logic in the selection of transliterated representations in order to lessen the burden on those who wish to learn the system.

3. REFERENCES

ISO 646 - 1983 Information Processing - ISO 7 - bit coded character set for information interchange.

ISO 5426 - 1983 Extension of the Latin alphabet coded character set for bibliographic information interchange.

Notification of the Royal Institute Concerning the Transcription of Thai Characters into Roman. (1954)

TIS 620 - 2529 (1986) Thai Character Codes for Computer Application.

4. DEFINITIONS

4.1 "Characters" means alphabetic letters, digits, special markers, and other markers.

4.2 "Thai Characters" means Thai alphabetic letters, Thai digits, and special markers which are divided into the following:

4.2.1 Thai alphabetic letters;

4.2.1.1 Consonants:

ก ข ฃ ค ฅ ฆ ง จ ฉ ช ซ ฌ ญ ฎ ฏ ฐ ฑ ฒ ณ

ด ต ถ ท ธ น บ ป ผ ฝ พ ฟ ภ ม ย ร ฤ ล ฦ ว

ศ ษ ส ห ฬ อ ฮ

4.2.1.2 Vowels

4.2.1.2.1 Vowels written on the same level as consonants:

ะ า ำ เ แ ไ ใ

4.2.1.2.2 Vowels written above the line (upper level characters):

ั ็ ิ ี ึ ื ุ ู

/. . .

4.2.1.2.3 Vowels written below the line (lower level characters):

ุ ู ฺ

4.2.1.3 Tone Marks and Cancellation Mark written in the uppermost level[1] (uppermost level characters):

่ ้ ๊ ๋ ์

4.2.2 Thai Digits:

๐ ๑ ๒ ๓ ๔ ๕ ๖ ๗ ๘ ๙

4.2.3 Special Markers:

ๆ ฯ ๏ ๚ ๛

4.3 "Special Symbols" refers to the four symbols used to differentiate one Roman character that represents the same sound converted from different Thai characters:

4.3.1 uppercore ‾ (ISO 5426 : 4/5)

4.3.2 underscore _ (ISO 5426 : 5/8)

4.3.3 underpoint . (ISO 5426 : 5/6)

4.3.4 horn ' (ISO 5426 : 4/14)

5. ARRANGEMENT OF CHARACTER SEQUENCE

5.1 The conversion is made character by character from left to right.

5.2 A consonant (characters # 1 - # 48 in 6.) can have either an upper level or a lower level character and/or an uppermost level character; the sequence of the characters must be as follows:

5.2.1 A consonant with either an upper level character or an uppermost level character:

- consonant
- upper level character or uppermost level character

5.2.2 A consonant with both an upper level character and an uppermost level character:

- consonant
- uppermost level character
- upper level character

5.2.3 A consonant with a lower level character:

- consonant
- lower level character

5.2.4 A consonant with an uppermost level character and a lower level character:

- consonant
- uppermost level character
- lower level character

It must be noted that both an upper level character and a lower level character cannot concurrently appear with the same consonant.

6. TYPING SEQUENCE

6.1 If a Roman character contains one special symbol, type the character before the symbol.

6.2 If a Roman character contains two special symbols, type the character first, then the upperscore, and the underpoint or the horn respectively.

7. ROMANIZED TRANSLITERATION TABLE

NO	THAI CHARACTER	ROMANIZED CHARACTER	CHARACTER CODE	COMMENTS
1	ก	k		Ko Kai
2	ข	k̄h		Kho Khai
3	ฃ	k̄h		Kho Khuat
4	ค	kh		Kho Khwai
5	ฅ	k'h		Kho Khon
6	ฆ	k̤h		Kho Rakhang
7	ง	ng		Ngo Ngu
8	จ	c		Co Can
9	ฉ	c̄h		Cho Ching
10	ช	ch		Cho Chang
11	ซ	s		So So
12	ฌ	c̤h		Cho Choe
13	ญ	ẏ		Yo Ying
14	ฎ	ḍ		Do Chada
15	ฏ	ṭ		To Patak
16	ฐ	ṯh		Tho Than
17	ฑ	t̲h		Tho Montho
18	ฒ	t'h		Tho Phu Thao
19	ณ	ṇ		No Nen
20	ด	d		Do Dek
21	ต	t		To Tao
22	ถ	t̄h		Tho Thung
23	ท	th		Tho Thahan

NO	THAI CHARACTER	ROMANIZED CHARACTER	CHARACTER CODE	COMMENTS
24	ธ	th̤		Tho Thong
25	น	n		No Nu
26	บ	b		Bo Bai Mai
27	ป	p		Po Pla
28	ผ	p̄h		Pho Phung
29	ฝ	f̄		Fo Fa
30	พ	ph		Pho Phan
31	ฟ	f		Fo Fan
32	ภ	ph̤		Pho Samphao
33	ม	m		Mo Ma
34	ย	y		Yo Yak
35	ร	r		Ro Rua
36	ฤ	v		Ro Ru
37	ฤๅ	z		Ro Ru (long)
38	ล	l		Lo Ling
39	ฦ	ɬ	ISO 5426: 7/8	Lo Lu
40	ฦๅ	q		Lo Lu (long)
41	ว	w		Wo Waen
42	ศ	s̤̄		So Sala
43	ษ	s̄́		So Rusi
44	ส	s̄		So Sua
45	ห	h̄		Ho Hip

/...

NO	THAI CHARACTER	ROMANIZED CHARACTER	CHARACTER CODE	COMMENTS
46	ฬ	l̤		Lo Chula
47	อ	x		O Ang
48	ฮ	h̤		Ho Nok Huk
49	ะ	a		Sara A
50	ั	a̤		Mai Han Akat
51	า	ā		Sara A (long)
52	ำ	ȧ		Sara Am
53	ิ	i		Sara I
54	ี	ī		Sara I (long)
55	ึ	ṳ		Sara U (unrounded)
56	ื	ṳ̄		Sara U (long)
57	ุ	u		Sara U (rounded)
58	ู	ū		Sara U (long rounded)
59	เ	e		Sara E
60	แ	æ	ISO 5426: 7/1	Sara Ae
61	โ	o		Sara O
62	ไ	ɪ	ISO 5426: 7/5	Sara Ai Mai Muan
63	ใ	i		Sara Ai Mai Malai
64	่	\	ISO 5426: 4/1	Mai Ek
65	้	^	ISO 5426: 4/3	Mai Tho

/...

NO	THAI CHARACTER	ROMANIZED CHARACTER	CHARACTER CODE	COMMENTS
66	◌่	◌́	ISO 5426: 4/2	Mai Tri
67	◌๋	◌̌	ISO 5426: 4/15	Mai Chattawa
68	◌์	◌̊	ISO 5426: 5/13	"Thanthakhat" silences a character or a group of characters when it appears above that character or the last character of the group such as สงฆ์ พักตร์ ชอล์ก ลักษมณ์.
69	◌๊	◌̆	ISO 5426: 4/6	"Mai Tai Khu" shortens a long vowel syllable.
70	ๆ	◌	ISO 5426: 3/8	"Mai Yamok" indicates a repetition of a word, a group of words or a phrase, or even a sentence.
71	ฯ	‡	ISO 5426: 3/6	"Paiyan Noi" indicates that the word has been shortened. It can combine with ล as ฯลฯ which is equivalent to etc., i.e.,et cetera.
72	◌็	◌̆	ISO 5426: 5/15	"Yamakkan" is written over the first member of a consonant cluster in old texts to indicate that it must be pronounced as a consonant cluster.

/...

NO	THAI CHARACTER	ROMANIZED CHARACTER	CHARACTER CODE	COMMENTS
73	๏	§	ISO 5426: 2/7	"Ta Kai" or "Fong Man" is used at the beginning of a stanza in poetry and also at the beginning of a paragraph in old texts. ๏" "Fong Man Fan Nu" occurs in older texts and can be considered a variant of ๏ .
74	๚	†	ISO 5426: 2/6	"Angkhan Dieo" is used at the end of a sentence, and at the end of a stanza in old texts. It is also used as a middle symbol in writing dates in the lunar calendar, for example 4 ๚ 15 6 meaning Wednesday, the 15th day (full moon) of the sixth lunar month. (Note: it looks very similar to ๆ , "Paiyan Noi" in # 71.)
75	๛	#	ISO 5426: 3/12	"Angkhan Khu" is used to terminate a paragraph in old texts. It can also combine with ฯ , "Kho Mut" in # 76 as ๛ฯ to terminate a story.

/...

NO	THAI CHARACTER	ROMANIZED CHARACTER	CHARACTER CODE	COMMENTS
76	๏	»	ISO 5426: 3/11	"Kho Mut" is written at the end of a story to indicate termination in old texts.
77	◌ं	◌	ISO 5426: 4/10	"Nikkhahit" is written above the rightmost shoulder of a single character to represent an 'ang' of 'am' sound, such as กังหัน pronouncing [săngkhan], and above the rightmost shoulder of a character and a vowel to represent an 'm' sound, such as จุ้ม pronouncing [chumnum].
78	◌ฺ	◌	ISO 5426: 5/4	"Phinthu" is written under a single consonant to signify a final or under the first member of double consonants to signify a consonant cluster in Pali texts.
79	๐	0		Sun
80	๑	1		Nung
81	๒	2		Song
82	๓	3		Sam
83	๔	4		Si
84	๕	5		Ha
85	๖	6		Hok

NO	THAI CHARACTER	ROMANIZED CHARACTER	CHARACTER CODE	COMMENTS
86	๗	7		Chet
87	๘	8		Pat
88	๙	9		Kao

ANNEX

CHARACTER AND SPECIAL SYMBOL SPECIFICATION

A1. Thai consonants with the same sound are represented by the same Roman character or a digraph of Roman characters. A special symbol or special symbols will be used in differentiating the Thai character equivalents, for example:

พ = ph

ภ = p̣h

A2. An "h" as a second member of a digraph signifies aspiration, for example:

ช = ch

พ = ph

A3. A special symbol is placed above, below, or at the right of a character of the first member of a digraph, for example:

ข = k̇h

ท = t̲h

ฒ = t'h

A4. High Class Consonants[2] are identified by an upperscore marker in order that automatic pronunciation rules can be easily derived from them, for example:

ฃ = k̄h

ศ = s̄

Remark : an example of an automatic pronunciation rule :

a consonant with an upperscore occurring before a long vowel with or without a sonorant will yield a rising tone.

A5. The less frequently used consonant of two different consonants with the same sound in the same consonant class (i.e. both Middle Class Consonants or both Low Class Consonants) is identified by underpoint, for example:

ฦ = l

ฬ = l̤

A6. A High Class Consonant is differentiated from a Low Class Consonant having the same sound by an upperscore, for example:

ฟ = f

ฝ = f̄

A7. If more than two different consonants have the same sound, a special symbol or symbols (as indicated in 3.3) must be employed. An underpoint, a horn, or an underscore will be used in descending order respectively, according to their frequency of occurrence in the Thai language, for example:

ท = th

ธ = t̤h

ฒ = t̛h

ฑ = t̲h

A8. Short and long vowels are specified by the same Romanized character. An upperscore mark identifies a long vowel, for example:

ิ = i

ี = ī

A9. The four Roman Characters q, v, x, and z are used in representing consonants with different phonetic value as follows:

ฦ with a phonetic value of [rɬ] for v.

ฦๆ with a phonetic value of [rɬɬ] for z.

ภๆ with a phonetic value of [lɬɬ] for q.

ฮ with a phonetic value of [ʔ] for x.

FOOTNOTES

[1]If a tone mark or a cancellation mark is directly over a consonant without an upper level character, it is adjusted to an upper level in typesetting and displaying on the computer screen for neatness of appearance.

[2]There are three consonant classes in the Thai language : High, Middle, and Low.

The High Class Consonants are ข ฃ ฉ ฐ ถ ผ ฝ ศ ษ ส ห. They are classified as High Class Consonants because a live syllable (a syllable ending with a long vowel or a sonorant) which has a High Class Consonant as its initial is pronounced with a rising tone and can take only ่ and ้ tone marks, yielding low and falling tones respectively; a dead syllable (a syllable ending in a short vowel or a stop consonant) which has a High Class Consonant as its initial is pronounced with a low tone and can take only ้ , yielding a falling tone.

The Middle Class Consonants are ก จ ฎ ฏ ฑ ฒ บ ป อ. They are classified as Middle Class Consonants because a live syllable which has a Middle Class Consonant as its initial is pronounced with a mid tone and can take ่ , ้ , ๋ and ๊ tone marks, yielding low, falling, high, and rising tones respectively; a dead syllable which has a Middle Class Consonant as its initial can take ้ , ๋ and ๊ tone marks, yielding falling, high, and rising tones respectively.

/...

The Low Class Consonants are ก ข ฃ ค ฅ ฆ ง จ ช ซ ฌ ญ ฑ ฒ ณ ท ธ น พ ฟ ภ ม ย ร ล ว ศ ษ ส ห ฬ อ ฮ. They are classified as Low Class Consonants because a live syllable which has a Low Class Consonant as its initial is pronounced with a mid tone and can take only ่ and ้ tone marks, yielding falling and high tones respectively; a dead syllable with a long vowel which has a Low Class Consonant as its initial is pronounced with a falling tone and can take ้ and ่ tone marks, yielding high and rising tones respectively; a dead syllable with a short vowel which has a Low Class Consonant as its initial is pronounced with a high tone and can take ่ and ่ tone marks, yielding falling and rising tones respectively.